Am. J. Hum. Genet. 60:717–728, 1997

HLA Class II DR-DQ Amino Acids and Insulin-Dependent Diabetes Mellitus: Application of the Haplotype Method

Ana M. Valdes, Shannon McWeeney, and Glenys Thomson

Department of Integrative Biology, University of California at Berkeley, Berkeley

Summary

Insulin-dependent diabetes mellitus (IDDM) HLA class II DRB1-DQA1-DQB1 data from four populations (Norwegian, Sardinian, Mexican American, and Taiwanese) have been analyzed to detect the amino acids involved in the disease process. The combination of sites DRB1#67 and 86; DQA1#47; and DQB1#9, 26, 57, and 70 predicts the IDDM component in these four populations, when the results and criteria of the haplotype method for amino acids, developed in the companion paper in this issue of the *Journal,* are used. The following sites, either individually, or in various combinations, previously have been suggested as IDDM components: DRB1#57, 70, 71, and 86; DQA1#52; and DQB1#13, 45, and 57 (DQB1#13 and 45 correlates 100% with DQB1#9 and 26). We propose that DQA1#47 is a better predictor of IDDM than is the previously suggested DQA1#52, and we add DRB1#67 and DQB1#70 to the HLA DR-DQ IDDM amino acids. We do not claim to have identified all HLA DR-DQ amino acids—or highly correlated sites—involved in IDDM. The frequencies and predisposing/protective effects of the haplotypes defined by these seven sites have been compared, and the effects on IDDM are consistent across the populations. The strongest susceptible effects came from haplotypes DRB1*0301/DQA1*0501/DQB1*0201 and DRB1*0401-5-7-8/DQA1*0301/DQB1*0302. The number of strong protective haplotypes observed was larger than the number of susceptible ones; some of the predisposing haplotypes were present in only one or two populations. Although the sites under consideration do not necessarily have a functional involvement in IDDM, they should be highly associated with such sites and should prove to be useful in risk assessment.

Received May 30, 1996; accepted for publication November 20, 1996.

Address for correspondence and reprints: Dr. Glenys Thomson, Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California at Berkeley, Berkeley, CA 94720-3140. E-mail: glenys@violet.berkeley.edu

0002-9297/97/6003-0029$02.00

Introduction

Insulin-dependent (type I) diabetes mellitus (IDDM) is the most common and serious chronic illness of childhood. In fact, IDDM is twice as common as all childhood cancers combined, and, until the discovery of insulin in 1922, the disease was always fatal (Winter and Atkinson 1992). IDDM is an autoimmune disease with multigene dependence; environmental factors also influence the disease process. Its pathogenesis is known to involve lymphocytic infiltration of the islets of Langerhans, culminating in the destruction of insulin-producing β-cells (Cahill and McDevitt 1981). Although multiple genes have been related to IDDM, the HLA-linked genes are the major susceptibility markers known to date, and, under a multiplicative model, constitute ~44% of the IDDM genetic component (for review, see, e.g., Todd 1995; Owerbach and Gabbay 1996; Thomson, in press).

Many HLA class II effects with IDDM have been reported, including various protective and predisposing DQ haplotypes as well as an increased risk for DR3/DR4 heterozygotes (for review, see, e.g., Svejgaard et al. 1980; Thomson 1988, 1995*a*; Thomson et al. 1988; Nepom 1990; Tait and Harrison 1991; Thorsby and Rønningen 1993; Cucca and Todd 1996). Further, genetic predisposition to IDDM involves multiple HLA loci possibly spanning HLA-A to DP (e.g., see Tait and Harrison 1991). High linkage disequilibrium in the HLA region, especially for the class II DR-DQ genes, has prevented precise identification of the susceptibility genes in IDDM. Nevertheless, linkage disequilibrium within the HLA complex can vary from one ethnic group to the next (Imanishi et al. 1992), as can disease incidence (Karvonen et al. 1993), making cross-ethnic group comparisons extremely valuable (Mijovic et al. 1991; Serjeantson and Easteal 1991).

It has been suggested that IDDM correlates with the function of HLA class II gene products, which, in susceptible individuals, may present self-peptides to autoreactive T-cells and may trigger the autoimmune aggression against pancreatic β-cells (Nepom 1990). Paradoxically, the autoreactive T-cell lines so far characterized in IDDM patients have been found to be restricted by class II molecules that are not those most commonly associated with the disease (Durinovic-Bello et al. 1994). An alternative model is that certain HLA molecules are as-

sociated with IDDM only because they are poorly protective (Sheehy 1992). No single model regarding major-histocompatibility-complex/peptide interactions is adequate to explain the various findings. The unequivocal identification, beyond ethnic lines, of those elements, within class II DR-DQ genes, that can account for the association with IDDM certainly will help clarify the functional role that HLA plays in this autoimmune disease.

Search for HLA Class II DR-DQ Amino Acids Involved in IDDM

We have analyzed four population groups—two with a very high incidence of IDDM (Norwegians, 17.6/100,000/year [Rønningen et al. 1991]; and Sardinians, 35/100,000/year [Cucca et al. 1993]), one with a moderate incidence of IDDM (Mexican Americans, 9.5/100,000/year [Erlich et al. 1993]), and one with a very low incidence of IDDM (Taiwanese, 1.35/100,000/year [Hu et al. 1993]).

Several HLA DQ alleles have been demonstrated to be strongly associated with IDDM (for review, see, e.g., Mijovic et al. 1991; Thorsby and Rønningen 1993). Initial molecular models of the HLA component in IDDM concentrated on the class II DQA1 and DQB1 genes (e.g., see Todd et al. 1989; Khalil et al. 1990; Sanjeevi et al. 1995). DQB1#57 (Asp protective) correlates with IDDM incidence (Dorman et al. 1990). The one-amino-acid model for susceptibility to IDDM (Todd et al. 1989) is, however, insufficient to explain the HLA component in IDDM, and the addition of DQA1#52 (Arg predisposing) with DQB1#57 was found to be more predictive (Khalil et al. 1990).

We have applied the haplotype method presented in the companion paper (Valdes and Thomson 1997) to identify, at the amino acid level, the HLA class II DR-DQ contribution to IDDM. Our initial analysis looked at all possible amino acid pairs (up to residue 90) within DQA1-DQB1, excluding sites that are 100% correlated; for example, DQA1#18, 45, 48, 56, 61, 64, 66, and 80 all segregate in a way identical to that of site 11, so we count all eight of them as only one site. (A list of 100%-correlated sites within DRB1-DQA1-DQB1 is provided in appendix A.) Excluding 100%-correlated sites left 39 sites and 741 pairs to be analyzed. Using the result presented in the companion paper (Valdes and Thomson 1997)—that is, that the haplotype method test statistic gives a closer fit to null expectations when some, compared with none, of the true predisposing factors are included—we kept the 32 (4%) of pairs whose mean and variance values of the resampled standardized χ^2 measure most resembled the expected values of 0 and 1 (see eq. [9] and text of Valdes and Thomson 1997).

We then examined all possible DQA1-DQB1 amino acid triplets (1,120 triplets) that included these pairs and identified those with the closest fit to the null hypothesis: all predisposing sites have been identified. On the basis of these results, we kept only triplets that included DQA1#47 and DQB1#70 and/or DQB1#57 and/or DQA1#52, and we investigated other amino acid combinations including these sites. The mean and variance of the resampled standardized χ^2 measure for some combinations of these DQA1-DQB1 sites are given in table 1.

Many DQ combinations of sites fall within the critical values previously defined ($-0.2 <$ mean < 0.2; $0.5 <$ variance < 1.8; Valdes and Thomson 1997) for one or all three of the population groups Mexican Americans, Norwegians, and Sardinians, and thus they point to the fundamental role that DQ plays in HLA predisposition to IDDM (table 1). It is important to bear in mind that differences in the distribution of alleles and haplotypes make some populations more informative than others. For example, in the Mexican American population, DQA1#52-DQB1#57 fits under the haplotype-method criteria for identifying predisposing sites. However, no combination of DQA1 and DQB1 sites fell within the critical values for all four populations. Furthermore, although many DQ combinations fall within the critical limit, the fit may improve with addition of other HLA-region sites (e.g., DRB1).

We did not observe any particular effect due to DQA1#69, which has been proposed, by some authors (Sanjeevi et al. 1995), to potentially play an important role in IDDM. Further, we should bear in mind that the sites that meet our criteria do not necessarily play a functional role in IDDM predisposition. The sites that we have picked may simply reflect their strong association with the actual sites involved. For example, one site that we have listed may be strongly associated with a pattern at three sites that together have a functional role in IDDM predisposition (see, e.g., Salamon et al. 1996). A case in point is sites DQB1#13 and DQB1#45. Sanjeevi et al. (1995) argue for a functional role of these sites. Yet, we do not see them making any difference in our test statistic. If we look carefully, however, at the patterns of amino acid distribution, all of the haplotypes distinguished by DQB1#13 and DQB1#45 are covered by sites DQB1#9 and DQB1#26.

Overall, DQA1#47 gave a better fit than DQA1#52, in many combinations. DQA1#47 and 52 divide the DQA1 alleles identically for DQA1*0101, 0102, 0103, and 0104, versus 0201; 69 does also. However, although DQA1#52 does not divide the remaining alleles, DQA1#47 distinguishes the DQA1 alleles 0301 and 0302 from 0401, 0501, and 0601. For these alleles, DQA1#69 distinguishes 0301, 0302, and 0501 from 0401 and 0601.

Evidence of DRB1 Involvement

In Caucasian, Mexican American, Chinese and Japanese populations, differential IDDM susceptibility

Table 1

Haplotype Method Applied to IDDM, Using Combinations of DQA1 and DQB1 Sites in Four Ethnic Groups

Site(s)		Mean (Variance)			
DQA1	DQB1	Mexican American	Norwegian	Sardinian	Taiwanese
52	57	.14 (1.06)	1.15[a] (2.19[b])	.38[a] (1.49)	2.94[a] (3.81[b])
47	57	.15 (1.24)	1.01[a] (2.01[b])	.48[a] (1.69)	1.84[a] (2.93[b])
52	9, 57	.17 (1.34)	1.11[a] (2.32[b])	.31[a] (1.27)	1.92[a] (2.58[b])
47	9, 57	.02 (.79)	.83[a] (1.76)	.39[a] (1.54)	1.09[a] (2.17[b])
52	9, 26, 57	.07 (.93)	1.09[a] (2.12[b])	.38[a] (1.47)	2.79[a] (3.00[b])
47	9, 26, 57	.03 (1.01)	.27[a] (.97)	.32[a] (1.36)	1.70[a] (2.67[b])
52	9, 26, 57, 70	.17 (.97)	1.09[a] (2.01[b])	.11 (.92)	2.67[a] (3.25[b])
47	9, 26, 57, 70	.18 (1.36)	.15 (.79)	.10 (1.08)	1.63[a] (2.50[b])
52	9, 26, 70	.23[a] (1.22)	.75[a] (1.71)	.57[a] (1.69)	4.07[a] (4.93[b])
47	9, 26, 70	.07 (1.11)	.03 (1.20)	.13 (.96)	2.32[a] (3.17[b])
47	9, 13, 26, 45, 57, 70	.19 (1.03)	.17 (1.45)	.14 (1.77)	1.63[a] (2.50[b])

[a] Value does not fall within the $-.2 < .2$ limits for resampled HLA data.
[b] Value does not fall within the $.5 < 1.8$ limits for resampled HLA data.

among individuals with identical DQA and DQB loci has shown that HLA-DRB1 also influences IDDM susceptibility (see e.g., Sheehy et al. 1989; Vicario et al. 1992; Cucca et al. 1993, 1995; Erlich et al. 1993; Huang et al. 1995; Tait et al. 1995; Harfouch-Hammond et al. 1996; Yasunaga et al. 1996). Both HLA DR and HLA DQ have similar functional roles as antigen-presenting molecules and may influence IDDM risk in a similar way (Cucca et al. 1995). Yasunaga et al. (1996) have documented a different contribution of DR and DQ in susceptibility and resistance to IDDM in Japanese and Norwegian patients. In their study, DQB1*0301 was negatively associated with IDDM, regardless of the associated DRB1 and DQA1 alleles on the haplotype, whereas DQB1*0302 was, overall, positively associated with IDDM. In the Japanese population, however, DRB1*0406 haplotypes are decreased in the patient population, even when they are associated with DQB1*0302. Furthermore, in Caucasian populations (Harfouch-Hammond et al. 1996), DRB1*0404 had a dominant protective effect, even when it was associated with DQB1*0302 and with the high-risk DR3 haplotype.

Sites from DRB1 were then added to the DQ sites in our analysis, to examine their role in predisposition to IDDM. We looked at some of the most interesting combinations from table 1, in conjunction with DRB1 polymorphic sites. We found that combinations {DRB1#86; DQA1#47 and 69; and DQB1#9, 26, 57, and 70} and {DRB1#67; DQA1#47; and DQB1#9, 26, 57, and 70} fell within the critical values in all four ethnic groups. However, the best fit was found with the combination {DRB1#67 and 86; DQA1#47; and DQB1#9, 26, 57, and 70} (table 2).

There might well be other DRB1 sites involved in IDDM, such as DRB1#71 (Ghabanbasani et al. 1994; Harfouch-Hammond et al. 1996) and, possibly, DRB1#70 (Harfouch-Hammond et al. 1996) and DRB1#57 (Awata et al. 1990); but their effect is not as dramatic or evident in the populations that we have studied. DRB1#71 seems to account for the major differences in binding specificity between DR4 alleles (Yamanaka et al. 1992; Hammer et al. 1995). Furthermore, the IDDM-protective/intermediate DRB1*0404 differs from the predisposing allele 0402 only at positions 70 and 71. The application of the haplotype method to other ethnic groups will prove very useful in discerning all DRB1 sites involved and in keeping track of variation, between ethnic groups, in linkage disequilibrium within the HLA region.

The amino acid sites selected for this study were found by looking first at DQA1-DQB1 and later at DRB1. If the order had been reversed, it is possible that different sites would have been found. However, these sites would have had the same informative value (i.e., ability to split haplotypes in the same way) as the ones described here. For example, DQA1#47 distinguishes the DQA1 alleles 0301 and 0302 from 0401, 0501, and 0601. Specifically, it divides DR3 (DQA1*0501) haplotypes from the DR4 (DQA1*0301) haplotypes. As tables 1 and 2 show, not all seven sites from the combination proposed are necessary in all populations, indicating that the informative value of a site depends on the particular setting of linkage disequilibrium among sites in a given population.

Predictive Value: Relationship between Incidence of IDDM and Frequency of Susceptible and Protective Haplotypes

Theory

If the IDDM effect of having a haplotype as defined by a number of polymorphic amino acid sites is consis-

Table 2

Haplotype Method Applied to IDDM, Using Combinations of DRB1, DQA1, and DQB1 Sites in Four Ethnic Groups

Site(s)			Mean (Variance)			
DRB1	DQA1	DQB1	Mexican American	Norwegian	Sardinian	Taiwanese
71	47, 69	9, 26, 57, 70	.17 (1.37)	.18 (.66)	.22[a] (1.49)	1.62[a] (1.81[b])
86	47	9, 26, 57, 70	.03 (1.67)	.20[a] (.82)	.07 (.98)	.14 (.77)
67	47	9, 26, 57, 70	−.03 (.61)	.12 (.86)	−.20[a] (.71)	.09 (.96)
71	47	9, 26, 57, 70	.22[a] (1.36)	.02 (.70)	.24[a] (1.67)	1.79[a] (1.83[b])
74	47	9, 26, 57, 70	.06 (1.36)	.13 (.89)	.25[a] (1.83[b])	1.78[a] (2.43[b])
85	47	9, 26, 57, 70	−.003 (.83)	.15 (1.04)	.16 (1.02)	.73[a] (1.39)
86	47, 69	9, 26, 57, 70	.16 (1.26)	.01 (.69)	−.01 (.89)	.03 (.72)
67	47, 69	9, 26, 57, 70	.05 (.77)	.19 (.87)	−.12 (1.14)	.17 (1.16)
67, 86	47	9, 26, 57, 70	.08 (.74)	.11 (.62)	−.15 (.63)	−.10 (.59)

NOTE.—Underlined values denote the combination with the best fit.
[a] Value does not fall within the $-.2 < .2$ limits for resampled HLA data.
[b] Value does not fall within the $.5 < 1.8$ limits for resampled HLA data.

tent for all ethnic groups, it is then desirable, for risk-assessment purposes, that susceptible haplotypes account for the vast majority of patient cases whereas protective or intermediate haplotypes account for the majority of controls. As we will prove, this is not always the case, because of the close relationship, overall, between disease prevalence and the frequencies of susceptible, intermediate, and protective haplotypes in patients and controls. Yet, the trends between prevalence and haplotype frequencies that we describe may prove useful in risk assessment if they are taken into account appropriately.

As a first step, we derive theoretical expectations, assuming a disease model for one locus with three alleles (haplotypes), of which one is predisposing (A_1), another intermediate (A_2), and the third protective (A_3). We denote by p_1, p_2, and p_3 their respective frequencies in the control population ($p_1 + p_2 + p_3 = 1$). The penetrance values for all six possible genotypes are defined as $A_1A_1 = s + w$, $A_1A_2 = s + k_1w$, $A_1A_3 = s + k_2w$, $A_2A_2 = s + k_3w$, $A_2A_3 = s + k_4w$, and $A_3A_3 = s$.
The prevalence of the disease in the population is given by

$$\begin{aligned} T &= p_1U + p_2V + p_3Z \\ &= s + w[p_1^2 + 2p_1(p_2k_1 + p_3k_2) \\ &\quad + p_2^2k_3 + 2p_2p_3k_4] , \end{aligned} \tag{1}$$

where

$$U = s + p_1w + p_2k_1w + p_3k_2w ;$$

$$V = s + p_1k_1w + p_2k_3w + p_3k_4w ;$$

$$Z = s + p_1k_2w + p_2k_4w ;$$

and s is the penetrance of sporadic cases (i.e., factors—genetic or environmental—other than HLA). Under the assumption that the HLA component of IDDM is the same across ethnic groups, the k_i's and w are constant. The allele (haplotype) frequencies among patients—q_1, q_2, and q_3—are given by

$$q_1 = \frac{p_1U}{T} ;$$

$$q_2 = \frac{p_2V}{T} ; \tag{2}$$

$$q_3 = \frac{p_3Z}{T} ;$$

$$p_1 = -\alpha_1 + \frac{\sqrt{4\alpha_1^2w^2 + 4w(T - s) - \alpha_2}}{2w} ; \tag{3a}$$

$$q_1 = \frac{[\sqrt{4\alpha_1^2w^2 + 4w(T - s) - \alpha_2} - 2\alpha_1w][s + \sqrt{w(T - s) + w^2\alpha_3}]}{2Tw} ; \tag{3b}$$

$$q_2 = \left(\frac{p_2}{T}\right)\left\{[-2\alpha_1w + \sqrt{4\alpha_1^2w^2 + 4w(T - s) - \alpha_2}]\,\frac{k_1}{2} + s + w(k_3p_2 + k_4p_3)\right\} ; \tag{3c}$$

where

$$\alpha_1 = p_2k_1 + p_3k_2 ;$$

$$\alpha_2 = 4k_3p_2^2w^2 + 8p_2p_3k_4 ;$$

$$\alpha_3 = \alpha_1^2 - k_3p_2 + 2p_2p_3(k_1k_2 - k_4) .$$

Table 3

Disease-Model Parameters Used in Figure 1

Parameter	Model 1	Model 2	Model 3
w	.05	.05	.05
k_1	.8	1	.1
k_2	.4	.1	.05
k_3	.4	.05	.01
k_4	.2	.01	.01
s	.005	.005	.005

We evaluated p_1, q_1, q_2, and T (simultaneously) under three different disease models (table 3), using the above equations. First, we kept the frequency of intermediate haplotypes among controls ($p_2 = .25$) constant. When the resulting values were plotted (fig. 1*A*), an increase in prevalence with increased frequency of predisposing haplotypes in the general population became evident (p_1). An increase in the frequency of susceptible haplotypes in patients (q_1) with increased prevalence was also observed for all three models (fig. 1*B*). In contrast, a decrease in the frequency of intermediate haplotypes (q_2) in patients accompanied increased prevalence (fig. 1*C*).

The decrease of intermediate haplotypes in patients is not a consequence of a lower frequency in controls; rather, when prevalence is high, a larger proportion of patient haplotypes are of the susceptible type. A lower prevalence is accompanied by fewer predisposing haplotypes in patients (fig. 1*B*), haplotypes whose place is taken by intermediate haplotypes. This pattern could give the erroneous impression that, in populations with lower disease prevalence, the contribution of sporadic cases is higher. That this is not the case is shown by figure 1*C*, where s has been kept constant.

We also explored the relationship between disease prevalence and the frequency of protective haplotypes in controls (p_3). It would be logical to expect a negative correlation, which certainly would be the case if an increase in protective haplotypes were accompanied by a decrease in susceptible ones. In order to explore the effect of protective haplotypes independently, we have fixed the frequency of susceptible haplotypes ($p_1 = .15$) in controls, in equation (3a). When T was plotted as a function of p_3, with p_1 kept constant, prevalence decreased with an increase in the frequency of protective haplotypes in controls, for models 1 and 2 (fig. 1*D*). For model 3, however, prevalence remained practically constant with changing p_3. This most likely is due to the lower penetrance of genotypes with A_3 in this model, compared with the other two models.

In figure 1, the slope of each curve (measured as the change in prevalence divided by the change in frequency) is shown, and it is interesting to note that, at least in the disease models presented here, the most marked trends relative to prevalence will be observed in the frequencies of predisposing and intermediate haplotypes in patients. On the other hand, our theoretical results predict a much smaller change in prevalence as a consequence of increased protective haplotypes.

IDDM Data

Theoretical expectations have been compared with the data at hand. To understand the differences among populations, we have grouped haplotypes according to the residues at sites DRB1#67 and 86; DQA1#47; and DQB1#9, 26, 57, and 70 (appendix B lists all the haplotypes defined by these seven sites in the four populations under study). With a focus on strong effects, haplotypes have been divided into three classes—putative susceptible, intermediate, and protective. We have classified as protective only those haplotypes (if present) whose frequency in controls is higher than that in patients, in all four ethnic groups; in at least one of the ethnic groups, the frequency among controls was three or more times higher than that in patients, and in at least one ethnic group it represented $\geq 5\%$ of all control haplotypes. Analogous criteria were used to classify susceptible haplotypes, and all other haplotypes were assigned to the intermediate class. The frequencies of susceptible, intermediate, and protective haplotypes, under the abovementioned criteria used in the four ethnic groups considered, are summarized in table 4.

In the absence of adequate prevalence data for the populations available, we have used incidence data. For a chronic disease such as diabetes, the prevalence rate greatly exceeds the incidence rate. It reflects not only the rate of appearance of diabetes in the population but how long the cases remain alive. Although, ideally, prevalence should be a linear function of incidence, in practice, social factors—such as the health care available to the population—affect prevalence; for example, a marked increase in the prevalence of IDDM followed the discovery of insulin (Krolewski and Warram 1985). Therefore, by using incidence rates (Karvonen et al. 1993) to make comparisons across ethnic groups, we avoid confounding genetic factors with social factors that are of no use for our purposes.

In the populations under study, we observed a significant positive correlation between the frequency of susceptible haplotypes, in both patients and controls, and disease incidence ($R^2 = .85$ and $R^2 = .98$, respectively; figs. 2*A* and *B*). A negative correlation between frequency of intermediate haplotypes among patients and disease incidence also was observed, as expected on the basis of our theoretical results ($R^2 = .99$; fig. 2*C*). We observed a weaker correlation between the frequency of protective haplotypes in controls and disease incidence ($R^2 = .69$; fig. 2*D*).

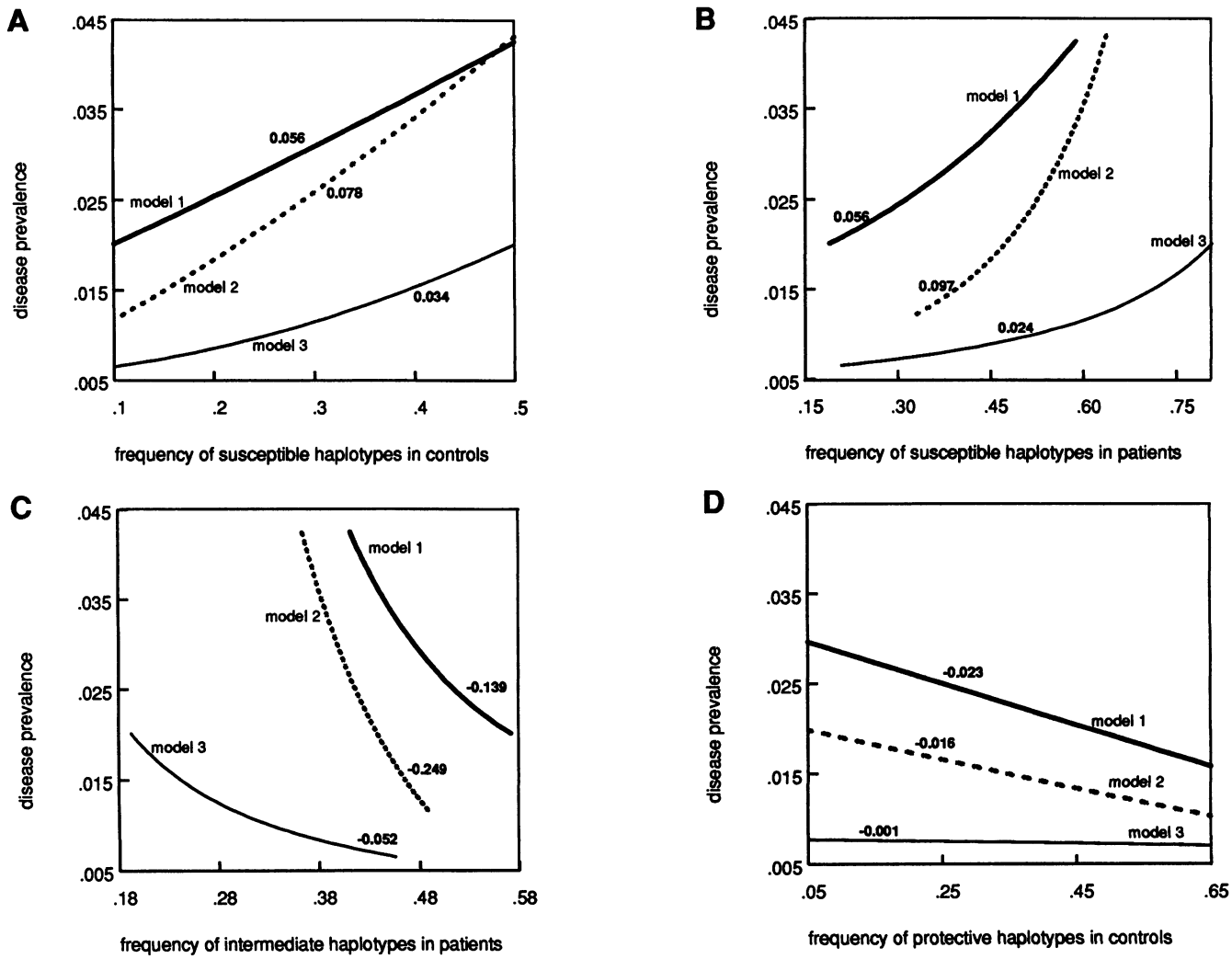

Figure 1 Theoretical relationship between prevalence and haplotype frequencies in patients and controls: prevalence vs. frequency of susceptible haplotypes in controls (p_1; eq. [3a]) (*A*), prevalence vs. frequency of susceptible haplotypes in patients (q_1; eq. [3b]) (*B*), prevalence vs. frequency of intermediate haplotypes in patients (q_2; eq. [3c]) (*C*), and prevalence vs. frequency of protective haplotypes in controls (p_3; eq. [3a], keeping p_1 constant) (*D*). The models used are given in table 3. The slope of each curve is indicated.

Discussion

The combination {DRB1#67 and 86; DQA1#47; and DQB1#9, 26, 57, and 70} predicts the predisposing component within the DR-DQ region in the Mexican American, Norwegian, Sardinian, and Taiwanese data considered here. We do not claim to have identified all HLA DR-DQ amino acids—or highly correlated sites—involved in IDDM. However, the sites identified are strong predictors of IDDM in the populations studied. Other ethnic groups are needed to define additional DRB1, DQA1, and DQB1 sites involved in IDDM.

It clearly has been demonstrated that other genes in the HLA region contribute to IDDM (e.g., see Tait and Harrison 1991; Robinson et al. 1993); for example, it is estimated that $<\sim 50\%$ of HLA DR3 haplotypes (which are basically homogeneous for the combination {DRB1*0301; DQA1*0501; DQB1*0201}) predispose to IDDM in Caucasians. This indicates that HLA-region variation additional to DRB1, DQA1, and DQB1 is required to define IDDM predisposition and/or protection. This variation could be within—or outside—the DR-DQ class II region. Variation in HLA-DPB1 has been implicated in IDDM (Erlich et al. 1996; Noble et al. 1996).

Although there are other HLA genes involved in IDDM, the results of this study reveal that a major factor is to be found within the DR-DQ region. Presumably, any other strong HLA-region IDDM components must not have high linkage disequilibrium with HLA DR-DQ. Moreover, only a few sites are needed to account for the IDDM component in the populations studied. We stress again that the combinations of sites presented here are valuable only for their predictive value in terms

Table 4

Frequency of Putative Protective, Intermediate, and Putative Susceptible Haplotypes in Four Ethnic Groups

DRB1 (Amino Acids)	DQA1 (Amino Acids)	DQB1 (Amino Acids)	Frequency in Controls/Frequency in Patients			
			Mexican American	Norwegian	Sardinian	Taiwanese
	Putative Protective					
1102/1201 (IV)	0501 (C)	0301 (nYDR)	.9/.0	3.0/.6	2.2/.0	13.3/.0
1501 (IV)	0102 (R)	0602 (FLDG)	2.3/.0	16.9/.6	.0/.0	3.6/1.1
0405/1402 (LG)	0501 (C)	0301 (nYDR)	10.5/3.5	.0/.0	3.8/.8	.0/.0
0701 (IG)	0201 (K)	0201 (YLAR)	7.8/5.8	9.1/1.7	4.6/.8	3.6/2.1
1101 (FG)	0501 (C)	0301 (nYDR)	5.0/.0	4.1/.0	8.8/.4	7.2/3.3
0801-2	0401/0501	0301				
1502 (IG)	0102-3 (R)	0601 (nYDR)	.5/.0	.0/.0	1.6/.0	16.4/2.1
0803	0102-3	0601/301				
Total			27.0/9.3	33.1/2.9	21.0/2.0	44.1/8.6
	Intermediate		54.8/43.1	45.1/34.4	50.6/17.9	50.8/52.3
	Putative Susceptible					
0401-5-7-8 (LG)	0301 (Q)	0302/201 (YLAR)	11.4/26.7	8.0/34.5	4.4/21.3	.0/6.5
0301 (LV)	0501 (C)	0201 (YLAR)	6.8/20.9	13.8/28.2	24.0/58.8	5.1/32.6
Total			18.2/47.6	21.8/62.7	28.4/80.1	5.1/39.1

of disease susceptibility. We cannot, at this time, surmise any functional involvement, in the disease process, of the molecular variants analyzed.

An important outcome of this study was the role of DQA1#47, rather than 52, in predicting disease predisposition at the haplotype level. DQA1#47 distinguishes the susceptible effect of DQA1*0301, which occurs on many DR4 haplotypes, from that of other IDDM-susceptibility DQA1 alleles, such as *0501, which occurs on DR3 haplotypes. In other studies, DRB1#86 has been discussed as being involved in IDDM (Hamaguchi et al. 1992; Erlich et al. 1993). To our knowledge, DRB1#67 has not been implicated previously in IDDM. Also, in other studies, positions DQB1#57 and DQB1#13 and 45 (which cosegregate identically with 9 and 26) have been proposed as being important in IDDM (Todd et al. 1989; Sanjeevi et al. 1995). As discussed earlier, variation at DRB1#71 (Ghabanbasani et al. 1994; Harfouch-Hammond et al. 1996) and, possibly, 70 (Harfouch-Hammond et al. 1996) and DRB1 57 (Awata et al. 1990) was found not to be important in the populations in this study but may be detected in study of other population groups.

The fact that DRB1#71 was not relevant in our analysis might be related to the fact that the frequency of DRB1*0402 is $\leq$2.7% in all of the control populations that we considered. We suspect that, if allele *0402 had been present at a higher frequency in the populations that we studied, we would have needed to include 71 and/or 70.

The classification of haplotypes that uses the combination {DRB1#67 and 86; DQA1#47; and DQB1#9, 26, 57, and 70} leads to the observation of consistent predisposing/protective effects across ethnic groups. The four populations studied exhibit different haplotypes, some of them unique, and the haplotype frequencies overall are remarkably different, so it is very difficult to find a common trend. Defining haplotypes by using only seven sites, we obtained a reduced number of haplotypes and consistent effects on all four populations. By analyzing the data within a mathematical framework, we were able to tackle the complexity of this multifactorial disease, in spite of the huge variation in genetic composition.

It must be noted that, although the sites presented in this study clearly correlate with disease predisposition, because disease operates at the genotypic level, these sites may or may not be able to predict genotypic risks for IDDM (Clerget-Darpoux et al. 1991). Nevertheless, the work developed at the haplotype level lays the ground for studies at the genotypic level, since it discards many possibilities.

The disease model presented in this paper is very general and makes no assumptions regarding mode of inheritance. The basic assumption made is that haplotypes can be divided into three major classes according to their effect on the disease process and that, within classes of haplotypes, mode of inheritance does not vary in a significant way, as would be the case if, for instance, some haplotypes predisposed to the disease in an overdominant way whereas others did so in a recessive way. Until evidence

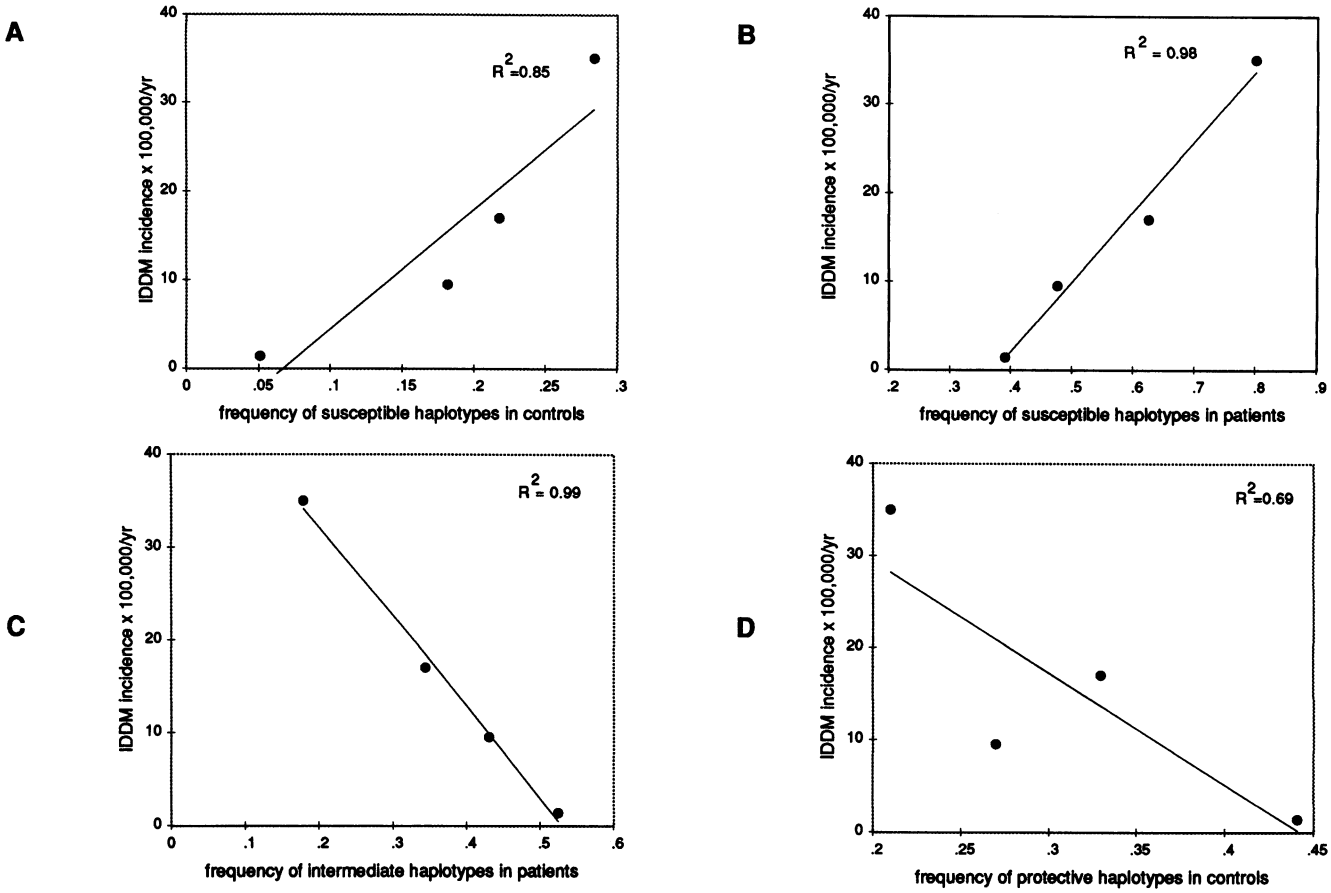

Figure 2 Observed relationship between IDDM incidence and haplotype frequencies in patients and controls in Sardinian, Norwegian, Mexican American, and Taiwanese populations: incidence vs. frequency of susceptible haplotypes in controls (R^2 = .85) (*A*), incidence vs. frequency of susceptible haplotypes in patients (R^2 = .98) (*B*), incidence vs. frequency of intermediate haplotypes in patients (R^2 = .99) (*C*), and incidence vs. frequency of protective haplotypes in controls (R^2 = .69) (*D*).

of such phenomena is found for IDDM, the model presented here provides a useful framework for looking at the complexity of this disease. One relevant result derived from this model is that intermediate haplotypes will tend to appear as if they had a predisposing effect in populations with low disease incidences, which, in turn, translates to heterogenous risks when different populations are compared. Clearly, this is a factor to be taken into account in cross-ethnic disease studies.

The agreement between observed incidences of IDDM and the theoretical curves based on haplotype frequencies indicates that the basic assumption made by our mathematical model—that is, that the same mode of disease is operating beyond ethnic lines—does indeed hold true. This is remarkable, particularly if we consider, first, that we analyzed only data from four populations and, second, the large differences, in population structure, between these four groups. This result points to the fundamental relationship between haplotype frequencies and disease prevalence and stresses the importance of considering population prevalence/incidence in the estimation of risks presented by specific HLA alleles and haplotypes. The next step is to apply this method to a larger number of populations and to move beyond the haplotype level, to examine the interactions of amino acid sites at the genotype level (the authors' work in this area is in progress), including all aspects of the disease, with respect to modes of inheritance, genetic heterogeneity, affected-sib-pair data, etc. (for review, see Thomson 1991, 1995*a*, 1995*b*). We expect this kind of cross-ethnic study to increase greatly our ability to predict IDDM genetic risks in a consistent and reliable way, for both HLA components and non-HLA components.

Acknowledgments

This research was supported by NIH grant HD12731.

Appendix A

Table A1

100% Correlated Sites

Polymorphic Site	100% Correlated Sites	Polymorphic Site	100% Correlated Sites
DRB1 4	DRB1 14, 25	DQA1 11	DQA1 18, 45, 48, 56, 61, 64, 66, 80, 129, 218
DRB1 9			
DRB1 10		DQA1 25	
DRB1 11		DQA1 26	DQA1 55, 187
DRB1 12		DQA1 34	
DRB1 13		DQA1 40	DQA1 51
DRB1 16		DQA1 41	
DRB1 26		DQA1 47	
DRB1 28		DQA1 50	DQA1 53
DRB1 30		DQA1 52	
DRB1 31		DQA1 54	
DRB1 32		DQA1 69	
DRB1 33		DQA1 75	DQA1 156, 161, 163
DRB1 37		DQA1 76	
DRB1 38		DQA1 134	DQA1 207
DRB1 40	DRB1 166	DQA1 175	
DRB1 47		DQA1 215	
DRB1 57		DQB1 3	
DRB1 58		DQB1 9	
DRB1 60		DQB1 13	
DRB1 67		DQB1 14	
DRB1 70		DQB1 23	
DRB1 71		DQB1 26	
DRB1 73		DQB1 28	DQB1 46, 47, 52
DRB1 74		DQB1 30	
DRB1 77		DQB1 37	
DRB1 85		DQB1 38	
DRB1 86		DQB1 45	
DRB1 96		DQB1 53	DQB1 84, 85, 89, 90
DRB1 98	DRB1 104, 180	DQB1 55	
DRB1 107		DQB1 56	
DRB1 120		DQB1 57	
DRB1 133	DRB1 142	DQB1 66	DQB1 67
DRB1 140		DQB1 70	
DRB1 164		DQB1 71	
DRB1 181		DQB1 74	
DRB1 233		DQB1 75	
		DQB1 77	
		DQB1 86	
		DQB1 87	

Appendix B

DRB1 #67, 86	DQA1 #47	DQB1 #9, 26, 57, 70	DRB1	DQA1	DQB1	DRB1 #67, 86	DQA1 #47	DQB1 #9, 26, 57, 70	DRB1	DQA1	DQB1
FG	C	FGDE	0801	0401	0401	IV	R	YGSG	1201	0102	0502
			0801	0401	0402				1501	0102	0502
			0802	0401	0402				1501	0103	0502
FG	C	FLDG	1101	0501	0602	IV	R	YGVG	1501	0101	0501
FG	C	nYDR	0801	0401	0301	IV	R	YLDG	1301	0103	0603
			0802	0501	0301				1501	0102	0603
			1101	0501	0301		C	FGDE	0302	0401	0402
FG	Q	nYDR	0901	0301	0301	LG	C	nYDR	1402	0501	0301
FG	Q	YLAR	0901	0301	0201				0405	0501	0301
FG	Q	YLDR	0901	0301	0303				1602	0501	0301
FG	R	nYDR	1601	0102	0601	LG	Q	nYDR	0401	0301	0301
FG	R	YGSG	1101	0102	0502				0405	0301	0301
			1601	0102	0502				0407	0301	0301
FV	C	FGDE	0804	0401	0402				0408	0301	0301
FV	C	nYDR	1103	0501	0301	LG	Q	YLAR	0401	0301	0302
			1104	0501	0301				0405	0301	0201
FV	C	YLDG	1104	0501	0603				0405	0301	0302
IG	C	nYDR	1303	0501	0301				0407	0301	0302
IG	C	YLVR	1302	0501	0604				0408	0301	0302
IG	K	YLAR	0701	0201	0201				0410	0301	0302
IG	K	YLDR	0701	0201	0303	LG	R	nYDR	1402	0101	0301
IG	Q	YLAR	1302	0301	0302	LG	R	YGVG	0101	0101	0501
IG	R	nYDR	0803	0103	0301				1001	0101	0501
			0803	0103	0601	LV	C	YGDG	1401	0501	0503
			1502	0102	0601	LV	C	YGSG	0301	0501	0502
			1502	0103	0601	LV	C	YLAR	0301	0501	0201
IG	R	YGVG	0103	0101	0501	LV	K	YLDR	1401	0201	0303
			1302	0101	0501	LV	Q	FGDE	0404	0301	0401
IG	R	YLVR	1302	0102	0604	LV	Q	nYDR	0403	0301	0301
			1302	0102	0605				0404	0301	0301
IV	C	nYDR	1102	0501	0301				0411	0301	0301
			1201	0501	0301	LV	Q	YLAR	0301	0301	0201
IV	C	YLDG	1102	0501	0603				0403	0301	0302
IV	Q	nYDR	0402	0301	0301				0404	0301	0302
IV	Q	YLAR	0402	0301	0302	LV	Q	YLDR	0404	0301	0303
IV	R	FLDG	1501	0102	0602	LV	R	FLDG	0301	0102	0602
IV	R	nYDR	1501	0102	0601	LV	R	YGDG	1401	0101	0503
IV	R	YGDG	1501	0101	0503	LV	R	YGSG	1401	0101	0502
						LV	R	YGVG	0102	0101	0501
									1401	0101	0501

References

Awata T, Kuzuya A, Matsuda A, Iwamoto Y, Kanazawa Y, Okuyama M, Juji T (1990) High frequency of aspartic acid at position 57 of HLA-DQβ-chain in Japanese IDDM patients and nondiabetic subjects. Diabetes 39:266–269

Cahill GF, McDevitt HO (1981) Insulin dependent diabetes mellitus: the initial lesion. N Engl J Med 304:1454–1465

Clerget-Darpoux F, Babron MC, Deschamps I, Hors J (1991) Complementation and maternal effect in insulin-dependent diabetes. Am J Hum Genet 49:42–48

Cucca F, Lampis R, Frau F, Masis D, Anguis E, Masile P, Chessa M, et al (1995) The distribution of DR4 haplotypes in Sardinia suggests a primary association of type I diabetes with DRB1 and DQB1 loci. Hum Immunol 43:301–308

Cucca F, Muntoni F, Lampis R, Frau F, Argiolas J, Silvetti M, Anguish E, et al (1993) Combinations of specific DRB1, DQA1, DQB1 haplotypes are associated with insulin-dependent diabetes mellitus in Sardinia. Hum Immunol 37:85–94

Cucca F, Todd JA (1996) HLA susceptibility to type I diabetes: methods and mechanisms. In: Browning MJ, McMichael AJ (eds) HLA and MHC: genes, molecules and function. BIOS Scientific, Oxford, pp 383–406

Dorman JS, LaPorte RE, Stone RA, Trucco M (1990) Worldwide differences in the incidence of type 1 diabetes are associated with amino acid variation at position 57 of the HLA-DQ beta chain. Proc Natl Acad Sci USA 87: 7370–7374

Durinovic-Bello I, Steinle A, Siegler AG, Schendel DJ (1994) HLA DQ restricted, islet-specific T-cell clones of a type I diabetic patient. Diabetes 43:1318–1325

Erlich H, Rotter JI, Chang J, Shaw S, Raffel LJ, Klitz W, Beshkov Y, et al (1993) HLA class II alleles and susceptibility to insulin-dependent diabetes mellitus in Mexican American families. Nat Genet 3:358–364

Erlich H, Rotter JI, Chang JD, Shaw SJ, Raffel LJ, Klitz W, Bugawan TL, et al (1996) Association of HLA-DPB1*0301 with IDDM in Mexican-Americans. Diabetes 45:610–614

Ghabanbasani MZ, Spaepen M, Buyse I, Marynen P, Bex M, Bouillon R, Cassiman JJ (1994) Improved risk assessment for insulin-dependent diabetes mellitus by analysis of amino acids in HLA-DQ and DRB1 loci. Eur J Hum Genet 2:177–184

Hamaguchi K, Kimura A, Dong RP, Noda N, Okeda T, Chikuba N, Nunoi K, et al (1992) Specific combinations of HLA-DRB and -DQB alleles confer susceptibility while the DQB gene determines resistance to insulin-dependent diabetes mellitus in Japanese. In: Tsuji K, Aizawa M, Sasazuki T (eds) Proceedings of the 11th International Histocompatibility Workshop and Conference. Vol 1: HLA 1991. Oxford University Press, Oxford, pp 488–492

Hammer J, Gallazzi F, Bono E, Karr R, Guenot M, Valsasnini P, Nagy Z, et al (1995) Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. J Exp Med 181:1847–1855

Harfouch-Hammond E, Timsit J, Boitard C, Bach J-F, Caillat-Zucman S (1996) Contribution of DRB1*04 variants to predisposition to or protection from insulin dependent diabetes mellitus is independent from DQ. J Autoimmun 9: 411–414

Hu C-Y, Allen M, Chuang L-M, Lin BJ, Gyllensten U (1993) Association of insulin-dependent diabetes mellitus in Taiwan with HLA class II DQB1 and DRB1 alleles. Hum Immunol 38:105–114

Huang H-S, Peng J-T, She JY, Zhang L-P, Chao CCK, Liu K-H, She J-X (1995) HLA-encoded susceptibility to insulin-dependent diabetes mellitus is determined by DR and DQ genes as well as their linkage disequilibria in a Chinese population. Hum Immunol 44:210–219

Imanishi T, Azaka T, Kimura A, Tokunaga K, Gojobori T (1992) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji K, Aizawa M, Sasazuki T (eds) Proceedings of the 11th International Histocompatibility Workshop and Conference. Vol 1: HLA 1991. Oxford University Press, Oxford, pp 1065–1220

Karvonen M, Tuomilehto J, Libman I, LaPorte R (1993) A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin dependent) diabetes mellitus. World Health Organization DIAMOND Project Group. Diabetologia 36:883–892

Khalil I, d'Auriol L, Gobet M, Morin L, Lepage V, Deschamps I, Park MS, et al (1990) A combination of HLA-DQβ Asp57-negative and HLA DQα Arg52 confers susceptibility to insulin-dependent diabetes mellitus. J Clin Invest 85:1315–1319

Krolewski AS, Warram JH (1985) Epidemiology of diabetes mellitus. In: Marble A, Krall LP, Bradley RF, Christlieb AR, Soeldner JS (eds) Joslin's diabetes mellitus. Lea & Fabger, Philadelphia, pp 12–42

Mijovic CG, Barnett AH, Todd JA (1991) Trans-racial gene mapping. In: Harrison LC, Tait BD (eds) The genetics of diabetes. Part 1. Bailliere Tindall, London, pp 321–340

Nepom GT (1990) A unified hypothesis for the complex genetics of HLA associations with IDDM. Diabetes 39:1153–1157

Noble J, Valdes AM, Cook M, Klitz W, Thomson G, Erlich H (1996) The role of HLA class II genes in insulin dependent diabetes mellitus (IDDM): molecular analysis of 180 Caucasian multiplex families. Am J Hum Genet 59:1134–1148

Owerbach D, Gabbay K (1996) The search for IDDM susceptibility genes: the next generation. Diabetes 45:544–551

Robinson WP, Thomson G, Barbosa J, Rich SS (1993) The homozygous parents affected sib pair method of detecting disease predisposition effects: application to insulin dependent diabetes mellitus. Genet Epidemiol 10:273–288

Rønningen KS, Spurkland A, Iwe T, Vartdal F, Thorsby E (1991) Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus. Tissue Antigens 37:105–111

Salamon H, Tarhio J, Rønningen K, Thomson G (1996) On distinguishing unique combinations in biological sequences. J Comp Biol 3:407–423

Sanjeevi CB, Lybrand TP, DeWeese C, Landin-Olsson M, Kochum I, Dahlquist G, Sundkvist G, et al (1995) Polymorphic amino acid variations in HLA-DQ are associated with systemic physical property changes and occurrence of IDDM. Diabetes 44:125–131

Serjeantson SW, Easteal S (1991) Cross-ethnic group comparisons of HLA class II alleles and insulin dependent diabetes mellitus. In: Harrison LC, Tait BD (eds) The genetics of diabetes. Part 1. Bailliere Tindall, London, pp 299–319

Sheehy MJ (1992) HLA and insulin-dependent diabetes: a protective perspective. Diabetes 41:123–129

Sheehy MJ, Scharf SJ, Rowe JR, Neme DE, Gimenez MH, Meske LM, Erlich HA, et al (1989) A diabetes-susceptible HLA haplotype is best defined by a combination of HLA-DR and -DQ alleles. J Clin Invest 83:830–835

Svejgaard A, Platz P, Ryder LP (1980) Insulin-dependent diabetes mellitus. In: Terasaki PI (ed) Histocompatibility 1980. University of California Press, Los Angeles, pp 638–656

Tait BD, Drummond BP, Varney MD, Harrison LC (1995) HLA-DRB1*0401 is associated with susceptibility to insulin-dependent diabetes mellitus independently of the DQB1 locus. Eur J Immunogenet 22:289–297

Tait BD, Harrison LC (1991) Overview: the major histocompatibility complex and insulin dependent diabetes mellitus. In: Harrison LC, Tait BD (eds) The genetics of diabetes. Part 1. Bailliere Tindall, London, pp 211–228

Thomson G (1988) HLA disease associations: models for insulin dependent diabetes mellitus and the study of complex human genetic disorders. Annu Rev Genet 22:31–50

——— (1991) HLA population genetics. In: Harrison LC, Tait BD (eds) The genetics of diabetes. Part 1. Bailliere Tindall, London, pp 247–260

——— (*1995a*) HLA disease associations: models for the study of complex human genetic disorders. Crit Rev Clin Lab Sci 32:183–219

——— (*1995b*) Analysis of complex human genetic traits: an ordered-notation method and new tests for mode of inheritance. Am J Hum Genet 57:474–486

——— Strategies involved in mapping diabetes genes: an overview. Diabetes Rev (in press)

Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald

MJ, Gottschall JL, Barbosa J, et al (1988) Genetic heterogeneity, modes of inheritance and risk estimates for a joint study of Caucasians with insulin dependent diabetes mellitus. Am J Hum Genet 43:799–816

Thorsby E, Rønningen KS (1993) Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 insulin dependent diabetes mellitus. Diabetologia 36:371–377

Todd J (1995) Genetic analysis of type I diabetes using whole genome approaches. Proc Natl Acad Sci USA 92:8560–8565

Todd JA, Aitman TJ, Cornall RJ, Ghosh S, Hall J, Hearne CM, Knight A, et al (1991) Genetic analysis of a complex multifactorial disease, autoimmune type 1 (insulin-dependent) diabetes. Res Immunol 142:483–501

Todd JA, Mijovic C, Fletcher J, Jenkins D, Bradwell AR, Barnett AH (1989) Identification of susceptibility loci for insulin dependent diabetes mellitus by trans-racial gene mapping. Nature 338:587–589

Valdes AM, Thomson G (1997) Detecting disease-predisposing variants: the haplotype method. Am J Hum Genet 59: 703–716 (in this issue)

Vicario JL, Martinez-Laso J, Correll A (1992) Comparison between HLA-DRB and DQ DNA sequences and classical serological markers as type I (insulin dependent) diabetes mellitus predictive risk markers in the Spanish population. Diabetologia 35:475–481

Winter WE, Atkinson MA (1992) Getting to the root of type I diabetes. Diabetes Forecast (May): 35–38

Yamanaka K, Kwok WW, Mickelson EM, Masewicz S, Nepom GT (1992) T cell receptor V beta selectivity in T cell lines alloreactive to HLA-Dw14. Hum Immunol 33:57–64

Yasunaga S, Kimura A, Hamaguchi K, Rønningen KS (1996) Different contribution of HLA-DR and -DQ genes in susceptibility and resistance to insulin dependent diabetes mellitus (IDDM) Tissue Antigens 47:37–48